Estimation of human signal detection performance from event-related potentials using feed-forward neural network model

Miloš Koska*, Roman Rosipal, Artur König
Institute of Measurement Science
Slovak Academy of Sciences
Dúbravská cesta 9, 842 19 Bratislava
Slovakia
Email: umerkosk, umerrosp, umerkoni@savba.sk

Leonard J. Trejo
Department of Psychology
University of Illinois at Urbana–Champaign
603 East Daniel Street, Champaign, IL 61820
USA
Email: trejo@uiuc.edu

Abstract We compared linear and neural network models for estimating human signal detection performance from event-related potentials (ERP) elicited by task-relevant stimuli. Data consisted of ERPs and performance measures from five trained operators who monitored a radar display and detected and classified visual symbols at three contrast levels. The performance measure (PF1) was a composite of accuracy, speed, and confidence of classification responses. The ERPs, which were elicited by the symbols, were represented in the interval 0-1500 ms post-stimulus at three midline electrodes (Fz, Cz, Pz) using either principal components (PCA) or coefficients of autoregressive (AR) models. We constructed individual models of PF1 from both PCA and AR representations using either linear regression or radial basis function (RBF) networks. Applying the normalized mean square error of approximation as a criterion, we found that the PCA representation was superior to AR and that RBF networks estimated PF1 much more accurately than linear regression. This suggests that nonlinear methods combined with suitable ERP feature extraction can provide more accurate and reliable estimates of display-monitoring performance than linear models.
Key words event-related potentials, performance, signal detection, RBF networks

* This work was partially supported by Slovak Grant Agency for Science (grants No. 2/2040/95 and 95/5305/468). MK was also supported by a stipend from BAMO, Austria.

1 Introduction

In many safety-critical applications (e.g., air traffic control, power plant operation, military applications) the control is based on the ability of human operators to detect and evaluate task-relevant signals in presented visual data. Performance quality of operators varies over time, often falling bellow acceptable limits, and may result in errors with serious consequences. The likelihood of such errors could be reduced if physiological methods for assessment of human performance were available.

ERPs reflect mental processes and are known to be related to human performance, including signal detection, confidence ratings, target identification and recognition, memory, tracking, and mental computation ([11], [12]). The utility of ERPs for performance assessment depends on a valid model for estimating the performance of the human operator . However, when we attempt to develop such a model we are confronted with the curse of dimensionality, which arises from the complexity of physiological data. For example, 225 data samples (dimensions) are required to describe a single 1.5 s segment of ERP data from three electrodes. To adequately sample the input space for this many dimensions, we would need many more observations than we can practically measure within a controlled experiment. In this contribution we attempt to overcome this problem using two projective data reduction techniques. We also combine data reduction with two types of regression models for estimating performance.

Prior research has shown that both autoregressive models (AR) ([5], [7], [13]) and principal components analysis (PCA) ([3], [12]) provide adequate reduction of ERP data. Other research has shown that both linear regression and nonlinear neural networks can model the relationships between ERPs and performance [11]. However, so far the performance of nonlinear neural networks has not provided substantial improvement over linear models. In this study, we explore radial basis function networks (RBF) , which are known to efficiently map relationships of data exhibiting clusters [14], for ERP analysis.

2 Methods

2.1 Data sample construction

We have used ERPs acquired in an earlier study [11], during a signal detection task from five male Navy technicians experienced in the operation of display systems. Each technician was trained to a stable level of performance and tested in multiple blocks of 50–72 trials each on two separate days. Blocks were separated by 1-minute rest intervals. About 1000 trials were performed by each subject. Inter-trial intervals were of random duration with a mean of 3 s and a range of 2.5–3.5 s. The entire experiment was computer-controlled and performed with a 19-inch color CRT display (Figure 1).

Triangular symbols subtending 42 minutes of arc and of three different luminance contrasts (0.17, 0.43, or 0.53) were presented parafoveally at a constant eccentricity of 2 degrees visual angle. One symbol was designated as the target, the other as the non-target. On some blocks, targets contained a central dot whereas the non-targets did not. However, the association of symbols to targets was alternated between blocks to prevent the development of automatic processing. A single symbol was presented per trial, at a randomly selected position on a 2-degree annulus. Fixation was monitored with an infrared eye tracking device. Subjects were required to classify the symbols as targets or non-targets using button presses and then to indicate their subjective confidence on a 3-point scale using a 3-button mouse. Performance was measured as a linear composite of speed, accuracy, and confidence. A single measure, PF1, was derived using factor analysis of the performance data for all subjects, and validated within subjects. The computational formula for PF1 was

$$\text{PF1} = 0.33{*}\text{Accuracy} + 0.53{*}\text{Confidence} - 0.51{*}\text{Reaction Time}$$

using standard scores for accuracy, confidence, and reaction time based on the mean and variance of their distributions across all subjects. PF1 varied continuously, being high for fast, accurate, and confident responses and low for slow, inaccurate, and unconfident responses.

ERPs were recorded from midline frontal, central, and parietal electrodes (Fz, Cz, and Pz; [8]), referred to average mastoids, filtered digitally to a bandpass of 0.1 to 25 Hz, and decimated to a final sampling rate of 50 Hz. The prestimulus baseline (200 ms) was adjusted to zero to remove any DC offset. Vertical and horizontal electrooculograms (EOG) were also recorded. Epochs containing artifacts were rejected and EOG-contaminated epochs were corrected [6]. Furthermore, any trial in which no detection response or confidence rating was made by a subject was excluded along with the corresponding ERP.

Within each block of trials, a running-mean ERP was computed for each trial (Figure 2). Each running-mean ERP was the average of the ERPs over a window that included the current trial plus the 9 preceding trials for a maximum of 10 trials per average. Within this 10-trial window, a minimum of 7 artifact-free ERPs were required to compute the running-mean ERP. If fewer than 7 were available, the running mean for that trial was excluded. Thus each running mean was based on at least 7 but no more than 10 artifact-free ERPs. This 10-trial window corresponds to about 30 s of task time. The PF1 scores for each trial were also averaged using the same running-mean window applied to the ERPs, excluding PF1 scores for trials in which ERPs were rejected. Prior to analysis, the running-mean ERPs were clipped to extend from time zero (stimulus onset time) to 1500 ms post-stimulus, for a total of 75 time points.

2.2 Choice of regressors

The first step in model development is the choice of regressors. In our case the running-mean ERPs form the input variable $\mathbf{x}_i$ and running-mean performance factors PF1 form the output variable $f(\mathbf{x}_i)$, where $i = 1, 2, \ldots, N$ and N is the number of ERPs for a particular subject. The input variable $\mathbf{x}_i$ is represented by a vector of dimension 225 (75 time points for each of 3 electrodes). The output variable $f(\mathbf{x}_i)$ is represented by a scalar.

Since N gets values only between 400 and 900 (see Table 1, column ERPs) we are confronted with the curse of dimensionality for the approximation of the multidimensional function $f(\mathbf{x}_i)$. It was pointed out by Friedman and Stuetzle [4] that although a general solution to problem of approximating d-dimensional function is difficult, for many applications, good results can be obtained by observing that in the sparse estimation data $\{\mathbf{x}_i, f(\mathbf{x}_i)\}$, $\{\mathbf{x}_i\}$ are "concentrated" in one or a small number of regions with dimensions much less than d. In such cases the d-dimensional function, $f(\mathbf{x})$, can be approximated by

$$\hat{f}(\mathbf{x}) = \sum_{k=1}^{K} w_k f_k(\mathbf{P}_k \mathbf{x})$$

where $f_k(.)$ is a function defined for the kth region, w_k is its "weight" and $\mathbf{P}_k$ is the corresponding projection operator and a $d_1 \times d$ matrix, where $d_1 \ll d$.

In order to compare the two projection operators (PCA, AR) in combination with the two approximation schemes (linear regression, RBF network) the following steps have been taken:

1. For each electrode, covariance-based PCA was computed [1] and the 10 most significant factors were chosen. Typically, they accounted for 73 – 82 % of variance in the data. Factor scores were computed for each running-mean ERP and stored for model development.

2. ERPs from each electrode were parameterized by AR model of order 10. The order of the model was determined by Akaike's final prediction error criterion

$$FPE(l) = \log(\hat{\sigma}_l^2) + \frac{2l}{j}$$

   where l is model order, $\hat{\sigma}_l^2$ is the variance of residuals and j is the number of observations.

Both of the above transformations reduced the dimension of input vector $\mathbf{x}_i$ from 225 to 30.

2.3 RBF networks

Results of Zhang et al. [14] indicate that in the case of "concentrated" data good results may be achieved with RBF networks [1].

A RBF network with n inputs and a scalar output is given by

$$f_r(\mathbf{x}) = \sum_{i=1}^{m} \lambda_i \phi(\|\mathbf{x} - \mathbf{c}_i\|)$$

where $\mathbf{x} \in R^n$ is the input vector, $\phi(.)$ is a given basis function from R^+ to R, $\|.\|$ denotes the Euclidean norm, λ_i, $0 \le i \le m$, are the weights of basis functions, $\mathbf{c}_i \in R^n$, $1 \le i \le m$, are the RBF centers, and m is the number of basis functions. The advantage of RBF networks is that when the centers $\mathbf{c}_i$, their number, m, and the shape of the basis functions $\phi(.)$ are all fixed they can be viewed as a special case of linear regression model. Then it is possible to apply the orthogonal least squares algorithm for subset selection. The algorithm has the property that each selected center maximizes the increment of the explained variance of the approximated function and does not suffer numerical ill-conditioning problems [2].

[1] Although most applications of PCA to ERP data ignore electrode differences when deriving the covariance matrix, we found on average 23 % lower normalized mean square approximation error between our approach and the usual one.

2.4 Linear regression models

Performance of RBF networks was compared with ordinary linear regression models which have been constructed from PCA factors and AR coefficients using forward–selection stepwise approach (SAS PROC STEPWISE). F-ratio test significance level p for including and removing regressors was set to 0.15.

2.5 Model validation

Both models, RBF networks and linear regression models, were validated using 10-fold cross-validation, i.e. running-mean ERPs gathered for each subject were divided into 10 equally sized parts, and each part was used as a validation set for a models build on remaining data.

3 Results

Simulations were implemented in MATLAB using the package of routines provided by M. J. L. Orr [9]. In RBF networks the thin-plate spline function

$$\phi(\nu) = \nu^2 \log(\nu)$$

was used as the nonlinearity and direct links between input and output layers were included in order to easily capture the linear properties in regressions [10]. The quality of approximation was measured in terms of normalized mean square error

$$\mathrm{NMSE} = \frac{\sum_{i=1}^{N}\left(f(\mathbf{x}_i) - \hat{f}(\mathbf{x}_i)\right)^2}{\sum_{i=1}^{N}\left(f(\mathbf{x}_i) - \bar{f}(\mathbf{x}_i)\right)^2}$$

where

$$\bar{f}(\mathbf{x}_i) = \frac{1}{N}\sum_{i=1}^{N} f(\mathbf{x}_i)$$

and N is the number of ERPs for particular subject.

For the PCA representation, the RBF networks approximated PF1 with lower NMSE than linear models (Table 1). Across subjects the average NMSE for RBF networks was 0.1536, or 3.4 times lower than for linear models. For the AR representation, NMSE was generally higher than for PCA (Table 2). As for the PCA, however, RBF networks still yielded lower approximation error. Across subjects the NMSE for RBF networks was 0.3868, or 2.0 times lower than for linear models.

We also analyzed the NMSE in a two-way repeated measures analysis of variance, with representation (PCA, AR) and model type (linear, RBF) as factors. PCA resulted in significantly lower NMSE than AR across models (mean NMSE 0.34 vs. 0.58; $F[1,4] = 61.4,\ p < 0.001$). RBF networks resulted in significantly lower NMSE than linear models for both representations (mean NMSE 0.27 vs. 0.65; $F[1,4] = 68.0,\ p < 0.001$). Representation and model type appear to independently influence NMSE, as there was no evidence of a significant interaction. The best results were achieved with RBF networks using the PCA representation.

4 Discussion

In this paper we have compared approximation capabilities of ERP-based linear models and RBF network models for display-monitoring performance data. Using the NMSE as a criterion for model evaluation, we find that PCA representations of ERPs are superior to AR and RBF networks are superior to linear regression models. Both representation and model type independently combined to reduce NMSE, where the PCA-RBF combination provided the best estimation of PF1. Although these results were achieved at expense of higher computational costs, these costs are non-recurring in the sense that once the models are determined there is no cost difference in their application.

Prior research has not always shown an advantage of neural network versus linear models of human performance based on ERP data [12]. We note that the RBF networks developed here reduced NMSE by a factor of 2.4, with a mean of 0.27 over subjects and representations. In the best cases, NMSE was on the order of 0.10, which corresponds to about 90% of the variance being accounted for. Because the training and test set samples we used were taken from the blocks of trials to which a running mean process was applied, the serial correlations in the data could inflate the proportion of variance accounted for by the models. Future work with data collected in different sessions could serve to confirm the ability of such models to generalize to new data.

From a technical standpoint, future research should explore other, potentially more powerful, models and representations. For example, it would be interesting to apply the PPR wavelet networks proposed in [14], which may provide further decreases in approximation error. From a performance assessment standpoint it would interesting to apply some of these methods to other tasks, for example the running memory and mental computation tasks described by Trejo et al. [11], for which linear models provided better fits than for the signal detection task studied here.

5 Acknowledgments

We thank the Director of Training Research Department at the Navy Personnel Research and Development Center, San Diego for providing us with data. One of the authors (MK) would also like to thank K. Hornik (Department of Statistics, TU Wien) and G. Kubin (Department of Communications and Radio-Frequency Engineering, TU Wien) for valuable comments and providing computational resources.

References

[1] D. S. Broomhead and D. Lowe. Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2:321–355, 1988.

[2] S. Chen, C. F. N. Cowan, and P. M. Grant. Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Transactions on Neural Networks*, 2(2):302–309, 1991.

[3] E. Donchin. A multivariate approach to the analysis of averaged evoked potentials. *IEEE Transactions on Biomedical Engineering*, BME-13:131–139, 1966.

[4] J. H. Friedman and W. Stuetzle. Projection pursuit regression. *Journal of American Statistical Association*, 76(376):817–823, 1981.

[5] A. S. Gevins and A. Remond. *Method of analysis for brain electrical and magnetic signals. Handbook of electroencephalography and clinical neurophysiology*, volume 1. Elsevier, 1987.

[6] G. Gratton, M. G. Coles, and E. Donchin. A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, pages 468–484, 1983.

[7] B. H. Jansen, J. B. Bourne, and J. W. Ward. Autoregressive estimation of short segment spectra for computerized EEG analysis. *IEEE Transactions on Biomedical Engineering*, BME-28(9):630–638, 1981.

[8] H. Jasper. The ten-twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, pages 397–403, 1958.

[9] M. J. L. Orr. *MATLAB routines for subset selection and ridge regression in linear neural networks.* Centre of Cognitive Science, University of Edinburgh, 1996. http://www.cns.ed.ac.uk/people/mark.html.

[10] T. Poggio and F. Girosi. Networks for approximation and learning. *Proceedings of the IEEE*, 78(9):1481–1496, 1990.

[11] L. J. Trejo, A. F. Kramer, and J. Arnold. Event-related potentials as indices of display-monitoring performance. *Biological Psychology*, 40:33–71, 1995.

[12] L. J. Trejo and M. J. Shensa. Linear and neural network models for predicting human signal detection performance from event-related potentials: a comparison of the wavelet transform with other feature extraction methods. In *Proceedings of the Fith Workshop on Neural Networks: Academic / Industrial / NASA / Defense*, SPIE Volume 2204, pages 153–161, San Diego: Society for Computer Simulation, 1993.

[13] A. C. Tsoi, D. S. C. So, and A. Sergejew. Classification of electroencephalogram using ANN. In *Neural Information Processing Systems 6*, pages 1151–1158, 1994.

[14] J. Zhang., G. G. Walter, Y. Miao, and W. N. W. Lee. Wavelet neural networks for function learning. *IEEE Transactions on Signal Processing*, 43(6):1485–1497, 1995.

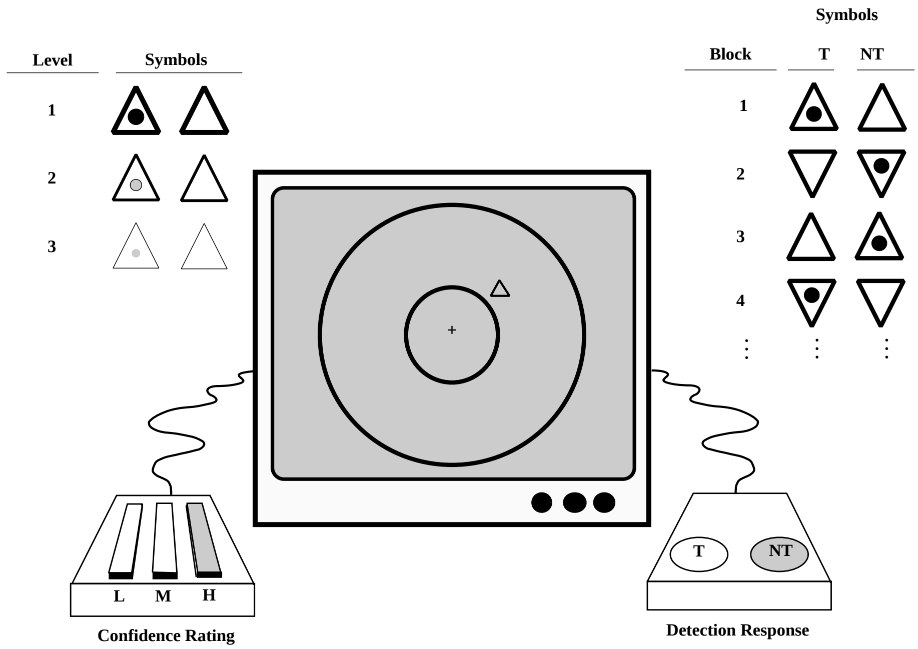

Figure 1: Display, input device configuration and symbols for task-relevant stimuli for the signal detection task.

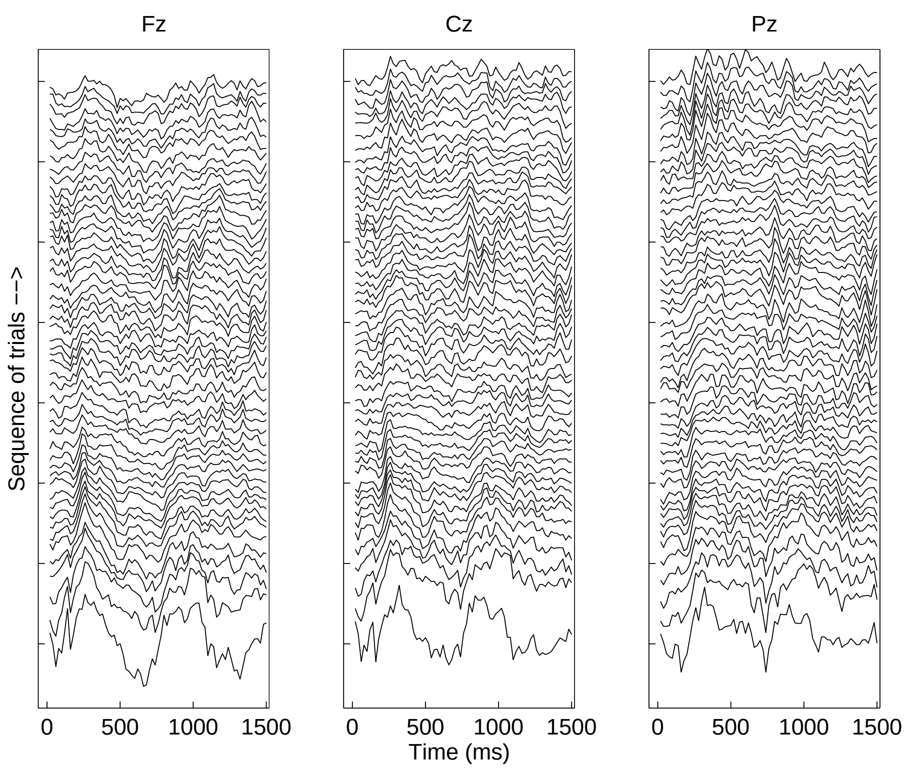

Figure 2: Running-mean ERPs at sites Fz, Cz and Pz for subject B in the first 50 running-mean ERPs.

Subj.	ERPs	Linear models				RBF networks			
		Model order		Test set NMSE		Hidden units		Test set NMSE	
		mean	(std)	mean	(std)	mean	(std)	mean	(std)
A	891	15	(3)	0.699	(0.029)	378	(56)	0.163	(0.030)
B	592	17	(3)	0.543	(0.046)	171	(42)	0.119	(0.028)
C	417	16	(1)	0.604	(0.113)	175	(26)	0.231	(0.050)
D	734	17	(2)	0.248	(0.027)	249	(109)	0.080	(0.020)
E	776	19	(2)	0.553	(0.037)	249	(60)	0.175	(0.025)

Table 1: Comparison of approximation errors (NMSE) achieved at test set with linear models and RBF networks using PCA representation. The values represent an average of 10 simulations with standard deviation in parentheses.

Subj.	ERPs	Linear models				RBF networks			
		Model order		Test set NMSE		Hidden units		Test set NMSE	
		mean	(std)	mean	(std)	mean	(std)	mean	(std)
A	891	12	(4)	0.836	(0.040)	436	(79)	0.378	(0.106)
B	592	18	(4)	0.694	(0.100)	260	(52)	0.384	(0.067)
C	417	6	(2)	0.799	(0.057)	197	(30)	0.454	(0.161)
D	734	14	(3)	0.657	(0.077)	373	(53)	0.372	(0.063)
E	776	6	(4)	0.861	(0.032)	286	(37)	0.346	(0.052)

Table 2: Comparison of approximation errors (NMSE) achieved at test set with linear models and RBF networks using AR representation. The values represent an average of 10 simulations with standard deviation in parentheses.